

# Generic minimizing behavior in semi-algebraic optimization

D. Drusvyatskiy\*      A.D. Ioffe†      A.S. Lewis‡

## Abstract

We present a theorem of Sard type for semi-algebraic set-valued mappings whose graphs have dimension no larger than that of their range space: the inverse of such a mapping admits a single-valued analytic localization around any pair in the graph, for a generic value parameter. This simple result yields a transparent and unified treatment of generic properties of semi-algebraic optimization problems: “typical” semi-algebraic problems have finitely many critical points, around each of which they admit a unique “active manifold” (analogue of an active set in nonlinear optimization); moreover, such critical points satisfy strict complementarity and second-order sufficient conditions for optimality are indeed necessary.

## 1 Introduction

Many problems of contemporary interest can broadly be phrased as an inverse problem: given a vector  $\bar{y}$  in  $\mathbf{R}^m$  find a point  $\bar{x}$  satisfying the inclusion

$$\bar{y} \in F(\bar{x}),$$

where  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  is some set-valued mapping (a mapping taking elements of  $\mathbf{R}^n$  to subsets of  $\mathbf{R}^m$ ) arising from the problem at hand. In other words, we would like to find a point  $\bar{x}$  such that the pair  $(\bar{x}, \bar{y})$  lies in the graph

$$\text{gph } F := \{(x, y) : y \in F(x)\}.$$

Stability analysis of such problems then revolves around understanding sensitivity of the solution set  $F^{-1}(\bar{y})$  near  $\bar{x}$  to small perturbations in  $\bar{y}$ . An extremely desirable property is for  $F$  to be *strongly regular* [40, Section 3G] at a pair  $(\bar{x}, \bar{y})$  in  $\text{gph } F$ , meaning that the graph of the inverse  $F^{-1}$  coincides locally around  $(\bar{y}, \bar{x})$  with the

---

\*Department of Mathematics, University of Washington, Seattle, WA 98195-4350; Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; <http://www.math.washington.edu/~ddrusv>; Research supported by AFOSR.

†Department of Mathematics, Technion-Israel Institute of Technology, Haifa, Israel 32000; [ioffe@math.technion.ac.il](mailto:ioffe@math.technion.ac.il). Research supported in part by US-Israel Binational Science Foundation Grant 2008261.

‡School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; <http://people.orie.cornell.edu/~aslewis/>. Research supported in part by National Science Foundation Grant DMS-0806057 and by the US-Israel Binational Scientific Foundation Grant 2008261.

graph of a single-valued Lipschitz continuous function  $g: \mathbf{R}^m \rightarrow \mathbf{R}^n$ . Naturally, then vectors  $\bar{y}$  for which there exists a solution  $\bar{x} \in F^{-1}(\bar{y})$  so that  $F$  is not strongly regular at  $(\bar{x}, \bar{y})$  are called *weak critical values* of  $F$ . We begin this work by asking the following question of Sard type:

Which mappings  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  have “almost no” weak critical values?

Little thought shows an immediate obstruction: the size of the graph of  $F$ . Clearly if  $\text{gph } F \subset \mathbf{R}^n \times \mathbf{R}^m$  has dimension (in some appropriate sense) larger than  $m$ , then no such result is possible. Hence, at the very least, we should insist that  $\text{gph } F$  is in some sense small in the ambient space  $\mathbf{R}^n \times \mathbf{R}^m$ .

Luckily, set-valued mappings having small graphs are common in optimization and variational analysis literature. Monotone operators make up a fundamental example: a mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^n$  is *monotone* if the inequality  $\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$  holds whenever the pairs  $(x_i, y_i)$  lie in  $\text{gph } F$ . Minty [34] famously showed that the graph of a maximal monotone mapping on  $\mathbf{R}^n$  is Lipschitz homeomorphic to  $\mathbf{R}^n$ , and hence monotone graphs can be considered small for our purposes. This property, for example, is fundamentally used in [37, 38]. The most important example of monotone mappings in optimization is the subdifferential  $\partial f$  of a convex function  $f$ . More generally, we may consider set-valued mapping arising from variational inequalities:

$$x \mapsto g(x) + N_Q(x),$$

where  $g$  is locally Lipschitz continuous and  $N_Q$  is the normal cone to a closed convex subset  $Q$  of  $\mathbf{R}^n$ . Such mappings appear naturally in perturbation theory for variational inequalities; see [40]. One can easily check that the graph of this mapping is locally Lipschitz homeomorphic to  $\text{gph } N_Q$ , and is therefore small in our understanding. In particular, we may look at conic optimization problems of the form

$$\min_x \{f(x) : G(x) \in K\},$$

for a smooth function  $f: \mathbf{R}^n \rightarrow \mathbf{R}$ , a smooth mapping  $G: \mathbf{R}^n \rightarrow \mathbf{R}^m$ , and a closed convex cone  $K$  in  $\mathbf{R}^m$ . Standard first order optimality conditions (under an appropriate qualification condition) amount to the variational inequality

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla f(x) + \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + N_{\{0\}^n \times K^*}(x, \lambda),$$

where  $K^*$  is the dual cone of  $K$  and the vector  $\lambda$  serves as a generalized Lagrange multiplier; see [40] for a discussion. Consequently the set-valued mapping on the right-hand-side again has a small graph.

In summary, set-valued mappings with small graphs appear often, and naturally so, in optimization problems. Somewhat surprisingly, assuming that the graph is small is by itself not enough to guarantee that strong regularity is typical — the conclusion that we seek. For instance, there exists a  $C^1$ -smooth convex function  $g: \mathbf{R} \rightarrow \mathbf{R}$  so that every number on the real line is a weakly critical value of the subdifferential  $\partial g$ . Such a function is easy to construct. Indeed, let  $f: \mathbf{R} \rightarrow \mathbf{R}$  be a surjective, continuous, and strictly increasing function whose derivative is zero almost everywhere (such a function  $f$  is described in [46] for example). Observe that

$f$  is nowhere locally Lipschitz continuous, since otherwise the fundamental theorem of calculus would imply that  $f$  is constant on some interval — a contradiction. On the other hand  $f$  is the derivative of the function  $h(t) := \int_0^t f(r) dr$ . The Fenchel conjugate  $h^*: \mathbf{R} \rightarrow \mathbf{R}$  is then exactly the function  $g$  that we seek. This example is interesting in light of Mignot's theorem [41, Theorem 9.65], which guarantees that at almost every subgradient, the inverse of the convex subdifferential must be single-valued and differentiable, though as we see, not necessarily locally Lipschitz continuous.

In light of this example, we see that even monotone variational inequalities can generically fail to be strongly regular. Incidentally, this explains the absence of Sard's theorem from all standard texts on variational inequalities (e.g. [19, 20, 40]), thereby deviating from classical mathematical analysis literature where implicit function theorems go hand in hand with Sard's theorem.

Motivated by optimization problems typically arising in practice, we consider *semi-algebraic* set-valued mappings — those whose graphs can be written as a finite union of sets each defined by finitely many polynomial inequalities. See for example [27] on the role of such mappings in nonsmooth optimization. In Theorem 3.7, we observe that any semi-algebraic mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ , whose graph has dimension no larger than  $m$ , has almost no weak critical values (in the sense of Lebesgue measure). Thus in the semi-algebraic setting, the size of the graph is the only obstruction to the Sard-type theorem that we seek.

Despite its simplicity, both in the statement and the proof, Theorem 3.7 leads to a transparent and unified treatment of generic properties of semi-algebraic optimization problems, covering in particular polynomial optimization problems, semi-definite programming, and copositive optimization — topics of contemporary interest. To illustrate, consider the family of optimization problems

$$\min_x f(x) + h(G(x) + y) - v^T x,$$

where  $f$  and  $h$  are semi-algebraic functions on  $\mathbf{R}^n$  and  $\mathbf{R}^m$ , respectively, and  $G: \mathbf{R}^n \rightarrow \mathbf{R}^m$  is a  $C^2$ -smooth semi-algebraic mapping. Here the vectors  $v, y$  serve as perturbation parameters. First order optimality conditions (under an appropriate qualification condition) then take the form of a generalized equation

$$\begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left( \partial f \times (\partial h)^{-1} \right)(x, \lambda),$$

where the subdifferentials  $\partial f$  and  $\partial h$  are meant in the limiting sense; see e.g. [41]. Observe that the perturbation parameters  $(v, y)$  appear in the range of the set-valued mapping on the right-hand-side. This set-valued mapping in turn, has a small graph. Indeed, the graphs of the subdifferential mappings  $\partial f$  and  $\partial h$  always have dimension exactly  $n$  and  $m$ , respectively [14, Theorem 3.7] (even locally around each of their points [11, Theorem 3.8], [12, Theorem 5.13]); monotonicity or convexity are irrelevant here. Thus the semi-algebraic Sard's theorem applies. In turn, appealing to some standard semi-algebraic techniques, we immediately conclude: for almost all parameters  $(v, y) \in \mathbf{R}^n \times \mathbf{R}^m$ , the problem admits finitely many composite critical points with each one satisfying a strict complementarity condition, a basic qualification condition (generalizing that of Mangasarian-Fromovitz) holds, both  $f$

and  $h$  admit unique active manifolds in the sense of [16, 30], and positivity of a second-derivative (of parabolic type) is both necessary and sufficient for second-order growth.

This development nicely unifies and complements a number of earlier results, such as the papers [44, 45] on generic optimality conditions in nonlinear programming, the study of the complementarity problem [42], generic strict complementarity and nondegeneracy in semi-definite programming [2, 43], as well as the general study of strict complementarity in convex optimization [13, 35]. In contrast, many of our arguments are entirely independent of the representation of the semi-algebraic optimization problem at hand. It is worth noting that convexity (and even Clarke regularity) is of no consequence for us. In particular, our results generalize and drastically simplify the main results of [3], where convexity of the semi-algebraic optimization problem plays a key role. Though we state our results for semi-algebraic problems, they all generalize to the “tame” setting; see [27] for the definitions. Key elements of the development we present here were first reported in [31]. In particular Theorem 7.3 in that work sketches the proof of generic minimizing behavior, restricted for simplicity to the case of linear optimization over closed semi-algebraic sets.

The outline of the manuscript is as follows. We begin in Section 2, by recording some basic notation to be used throughout the manuscript. In Section 3, we recall some rudimentary elements of semi-algebraic geometry and prove the semi-algebraic Sard theorem for weak critical values. In Section 4, we establish various critical point properties of generic semi-algebraic functions, while in Section 5, we refine the analysis of the previous section for semi-algebraic functions in composite form.

## 2 Basic notation

We begin by summarizing a few basic notions of variational and set-valued analysis. Unless otherwise stated, we follow the terminology and notation of [40, 41]. Throughout  $\mathbf{R}^n$  will denote an  $n$ -dimensional Euclidean space with inner-product  $\langle \cdot, \cdot \rangle$  and corresponding norm  $|\cdot|$ . We denote by  $B_\epsilon(x)$  an open ball of radius  $\epsilon$  around a point  $x$  in  $\mathbf{R}^n$ .

A set-valued mapping  $F$  from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ , denoted  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ , is a mapping taking points in  $\mathbf{R}^n$  to subsets of  $\mathbf{R}^m$ , with the *domain* and *graph* of  $F$  being

$$\text{dom } F := \{x \in \mathbf{R}^n : F(x) \neq \emptyset\},$$

$$\text{gph } F := \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^m : y \in F(x)\}.$$

We say that  $F$  is *finite-valued*, when the cardinality of the image  $F(x)$  is finite for every  $x \in \mathbf{R}^n$ .

A mapping  $\hat{F}: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  is a *localization* of  $F$  around  $(\bar{x}, \bar{y}) \in \text{gph } F$  if the graphs of  $F$  and  $\hat{F}$  coincide on a neighborhood of  $(\bar{x}, \bar{y})$ . The following is the central notion we explore.

**Definition 2.1** (Strong regularity and weak critical points). A set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  is  *$C^p$ -strongly regular* at  $(\bar{x}, \bar{y}) \in \text{gph } F$  if the inverse  $F^{-1}$  admits a  $C^p$ -smooth single-valued localization around  $(\bar{y}, \bar{x})$ .

A vector  $\bar{y} \in \mathbf{R}^m$  is a  *$C^p$ -weak critical value* of  $F$  if there exists a point  $\bar{x}$  in the preimage  $F^{-1}(\bar{y})$ , so that  $F$  is not  $C^p$ -strongly regular at  $(\bar{x}, \bar{y})$ .

Observe that  $\bar{y}$  being a weak critical value of  $F$ , at the very least, entails that the preimage  $F^{-1}(\bar{y})$  is nonempty. It is instructive to comment on the terms “strong” and “weak”. We use these to differentiate strong regularity from the weaker notion of metric regularity [25, 40] and the corresponding criticality concept. Note that the term “weakly critical” (with no qualifier) refers to the real-analytic version of the definition.

A mapping  $F: Q \rightarrow \tilde{Q}$ , where  $\tilde{Q}$  is a subset of  $\mathbf{R}^m$ , is  *$C^p$ -smooth* if for each point  $\bar{x} \in Q$ , there is a neighborhood  $U$  of  $\bar{x}$  and a  $C^p$ -smooth mapping  $\hat{F}: \mathbf{R}^n \rightarrow \mathbf{R}^m$  that agrees with  $F$  on  $Q \cap U$ . The symbol  $C^\omega$  will always mean real analytic. Smooth manifolds will play an important role in our work; a nice reference is [28].

**Definition 2.2** (Smooth manifolds). A subset  $\mathcal{M} \subset \mathbf{R}^n$ , is a  *$C^p$  manifold of dimension  $r$*  if for each point  $\bar{x} \in \mathcal{M}$ , there is an open neighborhood  $U$  around  $\bar{x}$  and a mapping  $F$  from  $\mathbf{R}^n$  to a  $(n - r)$ -dimensional Euclidean space so that  $F$  is  $C^p$ -smooth with the derivative  $\nabla F(\bar{x})$  having full rank and we have

$$\mathcal{M} \cap U = \{x \in U : F(x) = 0\}.$$

In this case, the *tangent space* to  $\mathcal{M}$  at  $\bar{x}$  is simply the set  $T_{\mathcal{M}}(\bar{x}) := \ker \nabla F(\bar{x})$ , while the *normal space* to  $\mathcal{M}$  at  $\bar{x}$  is defined by  $N_{\mathcal{M}}(\bar{x}) := \text{range } \nabla F(\bar{x})^*$ .

Given a  $C^1$ -smooth manifold  $\mathcal{M}$  and a mapping  $F$  that is  $C^1$ -smooth on  $\mathcal{M}$ , we will say that  $F$  has *constant rank on  $\mathcal{M}$*  if the rank of the operator  $\nabla \hat{F}(x)$  restricted to  $T_{\mathcal{M}}(x)$ , with  $\hat{F}$  being any  $C^1$ -smooth mapping agreeing with  $F$  on a neighborhood of  $x$  in  $\mathcal{M}$ , is the same for all  $x \in \mathcal{M}$ .

### 3 Semi-algebraic geometry and Sard’s theorem

Our current work is cast in the setting of semi-algebraic geometry. A *semi-algebraic* set  $Q \subset \mathbf{R}^n$  is a finite union of sets of the form

$$\{x \in \mathbf{R}^n : P_1(x) = 0, \dots, P_k(x) = 0, R_1(x) < 0, \dots, R_l(x) < 0\},$$

where  $P_1, \dots, P_k$  and  $R_1, \dots, R_l$  are polynomials in  $n$  variables. In other words,  $Q$  is a union of finitely many sets, each defined by finitely many polynomial equalities and inequalities. A map  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  is *semi-algebraic* if  $\text{gph } F \subset \mathbf{R}^{n+m}$  is a semi-algebraic set. For more details on semi-algebraic geometry, see for example [9, 47]. An important feature of semi-algebraic sets is that they can be decomposed into analytic manifolds. Imposing a very weak condition on the way the manifolds fit together, we arrive at the following notion.

**Definition 3.1** (Stratification). A  *$C^p$ -stratification* of a semi-algebraic set  $Q$  is a finite partition of  $Q$  into disjoint semi-algebraic  $C^p$  manifolds  $\{\mathcal{M}_i\}$  (called *strata*) with the property that for each index  $i$ , the intersection of the closure of  $\mathcal{M}_i$  with  $Q$  is the union of some  $\mathcal{M}_j$ ’s.

In particular, we can now define the dimension of any semi-algebraic set  $Q$ .

**Definition 3.2** (Dimension of semi-algebraic sets).

The *dimension* of a semi-algebraic set  $Q \subset \mathbf{R}^n$  is the maximal dimension of a semi-algebraic  $C^1$  manifold appearing in any  $C^1$ -stratification of  $Q$ .

It turns out that the dimension of a semi-algebraic set  $Q$  does not depend on any particular stratification. It is often useful to refine stratifications. Consequently, the following notation becomes convenient.

**Definition 3.3** (Compatibility). Given finite collections  $\{B_i\}$  and  $\{C_j\}$  of subsets of  $\mathbf{R}^n$ , we say that  $\{B_i\}$  is *compatible* with  $\{C_j\}$  if for all  $B_i$  and  $C_j$ , either  $B_i \cap C_j = \emptyset$  or  $B_i \subset C_j$ .

As we have alluded to at the onset, the following is a deep existence theorem for semi-algebraic stratifications [47, Theorem 4.8].

**Theorem 3.4** (Stratifications exist). *Consider a semi-algebraic set  $Q$  in  $\mathbf{R}^n$  and a semi-algebraic map  $F: Q \rightarrow \mathbf{R}^m$ . Let  $\mathcal{A}$  be a finite collection of semi-algebraic subsets of  $Q$  and  $\mathcal{B}$  a finite collection of semi-algebraic subsets of  $\mathbf{R}^m$ . Then there exists a  $C^\omega$ -stratification  $\mathcal{A}'$  of  $Q$  that is compatible with  $\mathcal{A}$  and a  $C^\omega$ -stratification  $\mathcal{B}'$  of  $\mathbf{R}^m$  compatible with  $\mathcal{B}$  such that for every stratum  $\mathcal{M} \in \mathcal{A}'$ , the restriction of  $F$  to  $\mathcal{M}$  is analytic and has constant rank, and the image  $F(\mathcal{M})$  is a stratum in  $\mathcal{B}'$ .*

Classically a set  $U \subset \mathbf{R}^n$  is said to be “generic”, if it is large in some precise mathematical sense, depending on context. Two popular choices are that of  $U$  being full-measure, meaning its complement has Lebesgue measure zero, and that of  $U$  being topologically generic, meaning it contains a countable intersection of dense open sets. In general, these notions are very different. However for semi-algebraic sets, the situation simplifies drastically. Indeed, if  $U \subset \mathbf{R}^n$  is a semi-algebraic set, then the following are equivalent.

- $U$  is dense.
- $U$  is full-measure.
- $U$  is topologically generic.
- The dimension of  $U^c$  is strictly smaller than  $n$ .

Complements of such sets are said to be *negligible*.

The following is the basic tool that we will use. A semi-algebraic finite-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  can be decomposed into finitely many  $C^\omega$ -smooth single-valued selections that “cross” almost nowhere. This result is standard: it readily follows for example from [14, Corollary 2.27]. We provide a proof sketch for completeness.

**Theorem 3.5** (Selections of finite-valued semi-algebraic mappings). *Consider a finite-valued semi-algebraic mapping  $G: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ . Then there exists an integer  $N$ , a finite collection of open semi-algebraic sets  $\{U_i\}_{i=0}^N$  in  $\mathbf{R}^n$ , and analytic semi-algebraic single-valued mappings*

$$G_i^j: U_i \rightarrow \mathbf{R}^m \quad \text{for } i = 0, \dots, N \text{ and } j = 1, \dots, i$$

*satisfying:*

1.  $\bigcup_i U_i$  is dense in  $\mathbf{R}^n$ .
2. For any  $x \in U_i$ , the image  $G(x)$  has cardinality  $i$ .
3. We have the representation

$$G(x) = \{G_i^j(x) : j = 1, 2, \dots, i\} \quad \text{whenever } x \in U_i.$$

*Proof.* Since  $G$  is semi-algebraic, there exists an integer  $N$  with the property that the cardinality of the images  $G(x)$  is no greater than  $N$  [47, Theorem 4.4]. For  $i = 0, \dots, k$ , define  $U_i$  to be the set of points  $x \in \mathbf{R}^n$  so that that image  $G(x)$  has cardinality precisely equal to  $i$ . A standard argument shows that the sets  $U_i$  are semi-algebraic. Stratifying, we replace each  $U_i$  with an open set (possibly empty) so that the union of  $U_i$  is dense in  $\mathbf{R}^n$ .

Fix now an index  $i$ . By [14, Corollary 2.27], there exists a dense open subset  $X_i$  of  $U_i$  with the property that there exists a semi-algebraic set  $Y_i \subset \mathbf{R}^m$  and a semi-algebraic homeomorphism  $\theta_i: \text{gph } G|_{X_i} \rightarrow X_i \times Y_i$  satisfying

$$\theta_i(\{x\} \times G(x)) = \{x\} \times Y_i \quad \text{for all } x \in X_i.$$

Observe that for each  $i$  the set  $Y_i$  has cardinality  $i$ . Enumerate the elements of  $Y_i$  by labeling  $Y_i = \{y_1, \dots, y_i\}$ . Define  $\pi$  to be the projection  $\pi(x, y) = y$  and for each  $j = 1, \dots, i$  set

$$G_i^j(x) = \pi \circ \theta_i^{-1}(x, y_j) \quad \text{for } x \in X_i.$$

Stratifying  $X_i$ , we may replace  $U_i$  by an open dense subset on which all the mappings  $G_i^j$  are analytic. The result follows.  $\square$

In particular, this theorem is applicable for semi-algebraic mappings with “small” graphs, since such mappings are finite-valued almost everywhere [14, Proposition 4.3].

**Theorem 3.6** (Finite selections for mappings with small graphs). *Suppose that the graph of a semi-algebraic set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  has dimension no larger than  $m$ . Then the inverse mapping  $F^{-1}: \mathbf{R}^m \rightrightarrows \mathbf{R}^n$  is finite-valued almost everywhere.*

We now arrive at the semi-algebraic Sard theorem — the main result of this section.

**Theorem 3.7** (Semi-algebraic Sard theorem for weakly critical values). *Consider a semi-algebraic set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  satisfying  $\dim \text{gph } F \leq m$ . Then the collection of weakly critical values of  $F$  is a negligible semi-algebraic set. More precisely, there exists an integer  $N$ , a finite collection of open semi-algebraic sets  $\{U_i\}_{i=0}^N$  in  $\mathbf{R}^n$ , and analytic semi-algebraic single-valued mappings*

$$G_i^j: U_i \rightarrow \mathbf{R}^m \quad \text{for } i = 0, \dots, N \text{ and } j = 1, \dots, i$$

*satisfying:*

1.  $\bigcup_i U_i$  is dense in  $\mathbf{R}^n$ .
2. For any  $x \in U_i$ , the preimage  $F^{-1}(x)$  has cardinality  $i$ .



3. We have the representation

$$F^{-1}(x) = \{G_i^j(x) : j = 1, 2, \dots, i\} \quad \text{whenever } x \in U_i.$$

*Proof.* Consider the open semi-algebraic sets  $\{U_i\}_{i=0}^N$  along with the single-valued, analytic, semi-algebraic mappings  $G_i^j : U_i \rightarrow \mathbf{R}^n$  provided by Theorems 3.5 and 3.6. Since for any  $y \in U_i$ , the preimage  $F^{-1}(y)$  has cardinality  $i$  and we have  $F^{-1}(y) = \{G_i^j(y) : j = 1, 2, \dots, i\}$ , we deduce that the values  $G_i^j(y)$  for  $j = 1, \dots, i$  are all distinct. Since  $G_i^j$  are in particular continuous, we deduce that the mapping  $F^{-1}$  has a single-valued analytic localization around  $(y, x)$  for every point  $x \in F^{-1}(y)$ . The result follows.  $\square$

We note that a Sard type theorem for semi-algebraic set-valued mapping with possibly large graphs, where criticality means absence of “metric regularity” [25, 40], was proved in [26]. Since we will not use this concept in the current work, we omit the details.

## 4 Critical points of generic semi-algebraic functions

In this section, we derive properties of critical points (appropriately defined) of semi-algebraic functions under generic linear perturbations. Throughout, we will consider functions  $f$  on  $\mathbf{R}^n$  taking values in the extended-real-line  $\overline{\mathbf{R}} = \mathbf{R} \cup \{+\infty\}$ . We will always assume that such functions are proper, meaning they are not identically equal to  $+\infty$ . The *domain* and *epigraph* of  $f$  are

$$\begin{aligned} \text{dom } f &:= \{x \in \mathbf{R}^n : f(x) < +\infty\}, \\ \text{epi } f &:= \{(x, r) \in \mathbf{R}^n \times \mathbf{R} : r \geq f(x)\}. \end{aligned}$$

The *indicator function* of a set  $Q \subset \mathbf{R}^n$ , denoted  $\delta_Q$ , is defined to be zero on  $Q$  and  $+\infty$  off it. A function  $f : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is *lower-semicontinuous* (lsc) whenever the epigraph  $\text{epi } f$  is closed. The notion of criticality we consider arises from the workhorse of variation analysis, the subdifferential.

**Definition 4.1** (Subdifferentials and critical points). Consider a function  $f : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and a point  $\bar{x}$  with  $f(\bar{x})$  finite.

1. The *proximal subdifferential* of  $f$  at  $\bar{x}$ , denoted  $\partial_p f(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  satisfying

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + O(|x - \bar{x}|^2).$$

2. The *limiting subdifferential* of  $f$  at  $\bar{x}$ , denoted  $\partial f(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  for which there exist sequences  $x_i \in \mathbf{R}^n$  and  $v_i \in \partial_p f(x_i)$  with  $(x_i, f(x_i), v_i)$  converging to  $(\bar{x}, f(\bar{x}), v)$ .
3. The *horizon subdifferential* of  $f$  at  $\bar{x}$ , denoted  $\partial^\infty f(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  for which there exist points  $x_i \in \mathbf{R}^n$ , vectors  $v_i \in \partial f(x_i)$ , and real numbers  $t_i \searrow 0$  with  $(x_i, f(x_i), t_i v_i)$  converging to  $(\bar{x}, f(\bar{x}), v)$ .



We say that  $\bar{x}$  is a *critical point* of  $f$  whenever the inclusion  $0 \in \partial f(\bar{x})$  holds.

The subdifferentials  $\partial_p f$  and  $\partial f$  generalize the notion of a gradient to the non-smooth setting. In particular, if  $f$  is  $C^2$ -smooth, then  $\partial_p f$  and  $\partial f$  simply coincide with the gradient  $\nabla f$ , while if  $f$  is convex, both subdifferentials coincide with the subdifferential of convex analysis [41, Proposition 8.12]. The horizon subdifferential  $\partial^\infty f$  plays an entirely different role: it detects horizontal normals to the epigraph of  $f$  and is instrumental in establishing calculus rules [41, Theorem 10.6]. For any set  $Q \subset \mathbf{R}^n$ , we define the *proximal* and *limiting normal cones* by the formulas  $N_Q^p := \partial_p \delta_Q$  and  $N_Q := \partial \delta_Q$ , respectively.

We will show in this section that any semi-algebraic function, subject to a generic linear perturbation, satisfies a number of desirable properties around any of its critical points. To this end, a key result for us will be that whenever  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is semi-algebraic, the graphs of the two subdifferentials  $\partial_p f$  and  $\partial f$  have dimension exactly  $n$  [14, Theorem 3.7]. (This remains true even in a local sense within the subdifferential graphs [11, Theorem 3.8], [12, Theorem 5.13]). Combining this with Theorem 3.7, we immediately deduce that generic subgradients of a semi-algebraic function are not weakly critical.

This observation, in turn, has immediate implications for minimizers of generic semi-algebraic functions, since strong regularity of the subdifferential is closely related to quadratic growth of the function. To be more precise, recall that  $\bar{x}$  is a *strong local minimizer* of a function  $f$  whenever there exist  $\alpha > 0$  and a neighborhood  $U$  of  $\bar{x}$  so that

$$f(x) \geq f(\bar{x}) + \frac{\alpha}{2}|x - \bar{x}|^2 \quad \text{for each } x \text{ in } U.$$

A more stable version of this condition follows.

**Definition 4.2** (Stable strong local minimizers). A point  $\bar{x}$  is a *stable strong local minimizer*<sup>1</sup> of a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  if there exist  $\alpha > 0$  and a neighborhood  $U$  of  $\bar{x}$  so that for every vector  $v$  near the origin, there is a point  $x_v$  (necessarily unique) in  $U$ , with  $x_0 = \bar{x}$ , so that in terms of the perturbed functions  $f_v := f(\cdot) - \langle v, \cdot \rangle$ , the inequality

$$f_v(x) \geq f_v(x_v) + \frac{\alpha}{2}|x - x_v|^2 \quad \text{holds for each } x \text{ in } U.$$

In [15, Proposition 3.1, Corollary 3.2], the authors show that strong metric regularity of the subdifferential at  $(x, v)$ , where  $x$  is a local minimizer of  $f_v := f(\cdot) - \langle v, \cdot \rangle$ , always implies that  $x$  is a stable strong local minimizer of  $f_v$ . See also [10, 17] for related results. Thus local minimizers of any semi-algebraic function, for a generic linear perturbation parameter, are stable strong local minimizers. Moreover, since the subdifferentials all have dimension exactly  $n$  and  $\dim(\text{gph } \partial f) \setminus (\text{gph } \partial_p f) \leq n$  it is easy to see that for a generic vector  $v$ , the strict complementarity condition

$$v \in \partial f(x) \implies v \in \text{ri } \partial_p f(x) \quad \text{holds for any } x \in \mathbf{R}^n.$$

We summarize all of these observations below.

---

<sup>1</sup>This notion appears under the name of uniform quadratic growth for tilt perturbations in [5], where it is considered in the context of optimization problems in composite form.

**Corollary 4.3** (Basic generic properties of semi-algebraic problems). *Consider an lsc, semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then there exists an integer  $N > 0$  such that for a generic vector  $v \in \mathbf{R}^n$  the function*

$$f_v(x) := f(x) - \langle v, x \rangle$$

*has no more than  $N$  critical points. In turn, each such critical point  $\bar{x}$  satisfies the strict complementarity condition*

$$0 \in \text{ri } \partial_p f_v(\bar{x}),$$

*and if moreover  $\bar{x}$  is a local minimizer of  $f_v$ , then  $\bar{x}$  is a stable strong local minimizer.*

We will see that by appealing further to semi-algebraic stratifications much more is true: any semi-algebraic function, up to a generic perturbation, admits a unique “stable active set”. To introduce this notion, we briefly record some notation. To this end, working with possibly discontinuous functions  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , it is useful to consider  $f$ -attentive convergence of a sequence  $x_i$  to a point  $\bar{x}$ , denoted  $x_i \xrightarrow{f} \bar{x}$ . In this notation

$$x_i \xrightarrow{f} \bar{x} \iff x_i \rightarrow \bar{x} \text{ and } f(x_i) \rightarrow f(\bar{x}).$$

An  $f$ -attentive localization of  $\partial f$  at  $(\bar{x}, \bar{v})$  is any mapping  $T: \mathbf{R}^n \rightrightarrows \mathbf{R}^n$  that coincides on an  $f$ -attentive neighborhood of  $\bar{x}$  with some localization of  $\partial f$  at  $(\bar{x}, \bar{v})$ .

It is often useful to require a kind of uniformity of subgradients. Recall that the subdifferential  $\partial f$  of an lsc convex function  $f$  is *monotone* in the sense that  $\langle v_1 - v_2, x_1 - x_2 \rangle \geq 0$  for any pairs  $(x_1, v_1)$  and  $(x_2, v_2)$  in  $\text{gph } \partial f$ . Relaxing this property slightly leads to the following concept [36, Definition 1.1].

**Definition 4.4** (Prox-regularity). An lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is called *prox-regular* at  $\bar{x}$  for  $\bar{v}$ , with  $\bar{v} \in \partial_p f(\bar{x})$ , if there exists a constant  $r > 0$  and an  $f$ -attentive localization  $T$  of  $\partial f$  around  $(\bar{x}, \bar{v})$  so that  $T + rI$  is monotone.

In particular  $C^2$ -smooth functions and lsc, convex functions are prox-regular at each of their points [41, Example 13.30, Proposition 13.34]. In contrast, the negative norm function  $x \mapsto -|x|$  is not prox-regular at the origin.

We are now ready to state what we mean by a “stable active set”. This notion introduced in [30], and rooted in even earlier manuscripts [1, 6–8, 18, 21, 22, 48], extends active sets in nonlinear programming far beyond the classical setting. The exact details of the definition will not be important for us, since we will immediately pass to an equivalent, but more convenient for our purposes, companion concept. Roughly speaking, a smooth manifold  $\mathcal{M}$  is said to be “active” or “partly smooth” for a function  $f$  whenever  $f$  varies smoothly along the manifold and sharply off it. The *parallel subspace* of any nonempty set  $Q$ , denoted  $\text{par } Q$ , is the affine hull of  $\text{conv } Q$  translated to contain the origin. We also adopt the convention  $\text{par } \emptyset = \emptyset$ .

**Definition 4.5** (Partial smoothness). Consider an lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and a  $C^p$  manifold  $\mathcal{M}$ . Then  $f$  is  $C^p$ -*partly smooth* ( $p \geq 2$ ) with respect to  $\mathcal{M}$  at  $\bar{x} \in \mathcal{M}$  for  $\bar{v} \in \partial f(\bar{x})$  if

1. **(smoothness)**  $f$  restricted to  $\mathcal{M}$  is  $C^p$ -smooth on a neighborhood of  $\bar{x}$ .
2. **(prox-regularity)**  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ .
3. **(sharpness)**  $\text{par } \partial_p f(\bar{x}) = N_{\mathcal{M}}(\bar{x})$ .
4. **(continuity)** There exists a neighborhood  $V$  of  $\bar{v}$ , such that the mapping,  $x \mapsto V \cap \partial f(\bar{x})$ , when restricted to  $\mathcal{M}$ , is inner-semicontinuous at  $\bar{x}$ .

In [16, Proposition 8.4], it was shown that the somewhat involved definition of partial smoothness can be captured more succinctly, assuming a strict complementarity condition  $\bar{v} \in \text{ri } \partial_p f(\bar{x})$ . Indeed, the essence of partial smoothness is in the fact that algorithms generating iterates, along with approximate criticality certificates, often “identify” a distinguished manifold in finitely many iterations; see the extensive discussions in [16, 24].

**Definition 4.6** (Identifiable manifolds). Consider an lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then a set  $\mathcal{M} \subset \mathbf{R}^n$  is a  $C^p$  identifiable manifold of  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  at a point  $\bar{x} \in \mathcal{M}$  for  $\bar{v} \in \partial f(\bar{x})$  if the set  $\mathcal{M}$  is a  $C^p$  manifold around  $\bar{x}$ , the restriction of  $f$  to  $\mathcal{M}$  is  $C^p$ -smooth around  $\bar{x}$ , and  $\mathcal{M}$  has the finite identification property: for any sequences  $x_i \xrightarrow{f} \bar{x}$  and  $v_i \rightarrow \bar{v}$ , with  $v_i \in \partial f(x_i)$ , the points  $x_i$  must lie in  $\mathcal{M}$  for all sufficiently large indices  $i$ .

In [16, Proposition 8.4], the authors showed that for  $p \geq 2$  the two sophisticated looking properties

1.  $f$  is  $C^p$ -partly smooth with respect to  $\mathcal{M}$  at  $\bar{x}$  for  $\bar{v}$ ,
2.  $\bar{v} \in \text{ri } \partial_p f(\bar{x})$ ,

taken together are simply equivalent to  $\mathcal{M}$  being a  $C^p$  identifiable manifold of  $f$  at  $\bar{x}$  for  $\bar{v} \in \partial_p f(\bar{x})$ . This will be the key observation that we will use with regard to partly smooth manifolds.

It is important to note that identifiable manifolds can fail to exist. For example, the function  $f(x, y) = (|x| + |y|)^2$  does not admit any identifiable manifold at the origin for the zero subgradient. On the other hand, we will see that such behavior, in a precise mathematical sense, is rare.

Roughly speaking, existence of an identifiable manifold at a critical point opens the door to Newton-type acceleration strategies [24, 29, 33] and moreover certifies that sensitivity analysis of the nonsmooth problem is in essence classical [23, 32]. To illustrate, we record two basic properties of identifiable manifolds [16, Propositions 5.9, 7.2], which we will use in Section 5.

**Theorem 4.7** (Basic properties of identifiable manifolds). *Consider an lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and suppose that  $\mathcal{M}$  is a  $C^2$ -identifiable manifold around  $\bar{x}$  for  $\bar{v} = 0 \in \partial_p f(\bar{x})$ . Then the following are equivalent*

1.  $\bar{x}$  is a strong local minimizer of  $f$ .
2.  $\bar{x}$  is a strong local minimizer of  $f + \delta_{\mathcal{M}}$ .

Moreover, equality

$$\text{gph } \partial f = \text{gph } \partial(f + \delta_{\mathcal{M}}),$$

holds on an  $f$ -attentive neighborhood of  $(\bar{x}, \bar{v})$ .

Generic existence of identifiable manifolds for semi-algebraic functions will now be a simple consequence of stratifiability of semi-algebraic sets. We note that, in particular, it shows that convexity is superfluous for the main results of [3].

**Corollary 4.8** (Generic properties of semi-algebraic problems). *Consider an lsc, semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then there exists an integer  $N > 0$  such that for a generic vector  $v \in \mathbf{R}^n$  the function*

$$f_v(x) := f(x) - \langle v, x \rangle$$

*has no more than  $N$  critical points. Moreover each such critical point  $\bar{x}$  satisfies*

1. **(prox-regularity)**  $f_v$  is prox-regular at  $\bar{x}$  for 0.
2. **(strict complementarity)** The inclusion  $0 \in \text{ri } \partial_p f_v(\bar{x})$  holds.
3. **(identifiable manifold)**  $f_v$  admits a  $C^\omega$  identifiable manifold at  $\bar{x}$  for 0.
4. **(smooth dependence of critical points)** The subdifferential  $\partial f$  is strongly regular at  $(\bar{x}, v)$ . More precisely, there exist neighborhoods  $U$  of  $\bar{x}$  and  $V$  of  $v$  so that the critical point mapping

$$w \mapsto U \cap (\partial f)^{-1}(w) = \{x \in U : x \text{ is critical for } f(\cdot) - \langle w, \cdot \rangle\}$$

*is single-valued and analytic on  $V$ , and maps  $V$  onto  $\mathcal{M}$ .*

*Moreover if  $\bar{x}$  is a local minimizer of  $f_v$  then  $\bar{x}$  is in fact a **stable strong local minimizer** of  $f_v$ .*

*Proof.* Generic finiteness of critical points and generic strict complementarity was already recorded in Corollary 4.3. We now tackle existence of identifiable manifolds. To this end, by [14, Theorem 3.7], the graph of the subdifferential mapping  $\partial f: \mathbf{R}^n \rightrightarrows \mathbf{R}^n$  has dimension  $n$ . Consequently, applying Theorem 3.7, we obtain a collection of open semi-algebraic sets  $\{U_i\}_{i=0}^k$  of  $\mathbf{R}^n$ , with dense union, and analytic semi-algebraic single-valued mappings

$$G_i^j: U_i \rightarrow \mathbf{R}^n \quad \text{for } i = 0, \dots, k \text{ and } j = 1, \dots, i$$

with the property that for each  $v \in U_i$  the set  $(\partial f)^{-1}(v)$  has cardinality  $i$  and we have the representation

$$(\partial f)^{-1}(v) = \{G_i^j(v) : j = 1, 2, \dots, i\}.$$

Let  $\mathcal{B}$  now be a stratification of  $\text{dom } f$  so that  $f$  is analytic on each stratum. Applying Theorem 3.4 to each  $G_i^j$ , we obtain a stratification  $\mathcal{A}_i^j$  of  $U_i$  so that  $G_i^j$  is analytic and has constant rank on each stratum of  $\mathcal{M}$  of  $\mathcal{A}_i^j$ , and so that  $f$  is analytic on the images  $G_i^j(\mathcal{M})$ . Finding a stratification of  $U_i$  compatible with  $\bigcup_j \mathcal{A}_i^j$ , we obtain a dense open subset  $\hat{U}_i$  of  $U_i$  so that around each point  $v \in \hat{U}_i$  there exists a neighborhood  $V$  of  $v$  so that  $G_i^j$  is analytic and has constant rank on  $V$ , and so that  $f$  is analytic on the images  $G_i^j(V)$ . Due to the constant rank condition, decreasing  $V$  further, we may be assured that  $G_i^j(V)$  are all analytic manifolds. Taking into account Theorem 3.7, we may also assume that none of the values in  $\hat{U}_i$

are weakly critical. Consequently for each  $v \in \hat{U}_i$ , there exists a sufficiently small neighborhood  $V$  of  $v$  so that the analytic manifold  $G_i^j(V)$  coincides with  $(\partial f)^{-1}(V)$  on a neighborhood of  $G_i^j(v)$ . Hence  $G_i^j(V)$  is an identifiable manifold at  $G_i^j(v)$  for  $v$ . Finally, appealing to Corollary 4.3, the result follows.  $\square$

Next we look more closely at second order growth, from the perspective of second derivatives. To this end, we record the following standard definition.

**Definition 4.9** (Subderivatives). Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and a point  $\bar{x}$  with  $f(\bar{x})$  finite. Then the *subderivative* of  $f$  at  $\bar{x}$  is defined by

$$df(\bar{x})(\bar{u}) := \liminf_{\substack{t \searrow 0 \\ u \rightarrow \bar{u}}} \frac{f(\bar{x} + tu) - f(\bar{x})}{t},$$

while for any vector  $\bar{v} \in \mathbf{R}^n$ , the *critical cone* of  $f$  at  $\bar{x}$  for  $\bar{v}$  is defined by

$$C_f(\bar{x}, \bar{v}) := \{u \in \mathbf{R}^n : \langle \bar{v}, u \rangle = df(\bar{x})(u)\}.$$

The *parabolic subderivative* of  $f$  at  $\bar{x}$  for  $\bar{u} \in \text{dom } df(\bar{x})$  with respect to  $\bar{w}$  is

$$d^2 f(\bar{x})(\bar{u}|\bar{w}) = \liminf_{\substack{t \searrow 0 \\ w \rightarrow \bar{w}}} \frac{f(\bar{x} + t\bar{u} + \frac{1}{2}t^2 w) - f(\bar{x}) - df(\bar{x})(\bar{u})}{\frac{1}{2}t^2}.$$

Some comments are in order. The directional subderivative  $df(\bar{x})(\bar{u})$  simply measures the rate of change of  $f$  in direction  $\bar{u}$ . Whenever  $f$  is locally Lipschitz continuous at  $\bar{x}$  we may set  $u = \bar{u}$  in the definition. The critical cone  $C_f(\bar{x}, \bar{v})$  denotes the set of directions  $u$  along which the directional derivative at  $\bar{x}$  of the function  $x \mapsto f(x) - \langle \bar{v}, x \rangle$  vanishes. The parabolic subderivative  $d^2 f(\bar{x})(\bar{u}|\bar{w})$  measures the second order variation of  $f$  along points lying on a parabolic arc, and hence the name. In particular, when  $f$  is  $C^2$  smooth at  $\bar{x}$ , we have

$$d^2 f(\bar{x})(\bar{u}|\bar{w}) = \langle \nabla^2 f(\bar{x})\bar{u}, \bar{u} \rangle + \langle \nabla f(\bar{x}), \bar{w} \rangle.$$

This three constructions figure prominently in second-order optimality conditions. Namely, if  $\bar{x}$  is a local minimizer of  $f$ , then  $df(\bar{x})(u) \geq 0$  for all  $u \in \mathbf{R}^n$ , and we have  $\inf_{w \in \mathbf{R}^n} d^2 f(\bar{x})(u|w) \geq 0$  for any nonzero  $u \in C_f(\bar{x}, 0)$ . On the other hand, deviating from the classical theory, the assumption  $df(\bar{x})(u) \geq 0$  for all  $u \in \mathbf{R}^n$  along with the positivity  $\inf_{w \in \mathbf{R}^n} d^2 f(\bar{x})(u|w) > 0$  for any nonzero  $u \in C_f(\bar{x}, 0)$ , guarantees that  $\bar{x}$  is a strong local minimizer of  $f$  only under additional regularity assumptions on the function  $f$ . See for example [5] or [41, Theorem 13.66] for more details.

We will now see that in the generic semi-algebraic set-up, the situation simplifies drastically: the parabolic subderivative completely characterizes quadratic growth at a critical point. The key to the development, not surprisingly, is the relationship between subderivatives of a function  $f$  and the subderivatives of the restriction of  $f$  to an identifiable manifold.

**Theorem 4.10** (First-order subderivatives and identifiable manifolds). *Consider an lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and suppose that  $f$  admits a  $C^2$  identifiable manifold  $\mathcal{M}$  at a point  $\bar{x}$  for  $\bar{v} \in \partial_p f(\bar{x})$ . Then for any  $u \in T_{\mathcal{M}}(\bar{x})$  we have*

$$df(\bar{x})(u) = d(f + \delta_{\mathcal{M}})(\bar{x})(u) = \langle \bar{v}, u \rangle.$$

*Proof.* Let  $g: \mathbf{R}^n \rightarrow \mathbf{R}$  be a  $C^2$ -smooth function coinciding with  $f$  on  $\mathcal{M}$  near  $\bar{x}$ . Standard subdifferential calculus implies

$$\partial_p f(\bar{x}) \subset \partial_p(f + \delta_{\mathcal{M}})(\bar{x}) = \partial_p(g + \delta_{\mathcal{M}})(\bar{x}) = \nabla g(\bar{x}) + N_{\mathcal{M}}(\bar{x}).$$

Moreover, one can easily verify  $d(g + \delta_{\mathcal{M}})(\bar{x})(u) = \langle \nabla g(\bar{x}), u \rangle$  for any  $u \in T_{\mathcal{M}}(\bar{x})$ . Since by the chain of inclusions above  $\bar{v}$  lies in  $\nabla g(\bar{x}) + N_{\mathcal{M}}(\bar{x})$ , we deduce

$$d(f + \delta_{\mathcal{M}})(\bar{x})(u) = d(g + \delta_{\mathcal{M}})(\bar{x})(u) = \langle \nabla g(\bar{x}), u \rangle = \langle \bar{v}, u \rangle.$$

Now since identifiable manifolds are partly smooth, we have  $\text{par } \partial_p f(\bar{x}) = N_{\mathcal{M}}(\bar{x})$ . Consequently we deduce

$$\text{aff } \partial_p f(\bar{x}) = \nabla g(\bar{x}) + N_{\mathcal{M}}(\bar{x}).$$

In particular, for any  $u \in T_{\mathcal{M}}(\bar{x})$  we have equality  $\langle \text{aff } \partial_p f(\bar{x}), u \rangle = \langle \nabla g(\bar{x}), u \rangle = \langle \bar{v}, u \rangle$ . On the other hand  $df(\bar{x})$  is the support function of the Fréchet subdifferential  $\hat{\partial}f(\bar{x})$  (see [41, Excercise 8.4]), and since  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ , we have  $\text{aff } \partial_p f(\bar{x}) = \text{aff } \hat{\partial}f(\bar{x})$ . We conclude  $df(\bar{x})(u) = \langle \bar{v}, u \rangle$ , as claimed.  $\square$

As a direct consequence, we deduce that critical cones are simply tangent spaces to identifiable manifolds, when the latter exist. The following is an extension of [14, Proposition 6.4].

**Theorem 4.11** (Critical cones and identifiable manifolds). *Consider an lsc function  $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$  and suppose that  $f$  admits a  $C^2$  identifiable manifold  $\mathcal{M}$  at a point  $\bar{x}$  for  $\bar{v} \in \partial_p f(\bar{x})$ . Then the critical cone coincides with the tangent space*

$$C_f(\bar{x}, \bar{v}) = T_{\mathcal{M}}(\bar{x}).$$

*Proof.* The inclusion  $C_f(\bar{x}, \bar{v}) \supset T_{\mathcal{M}}(\bar{x})$  is immediate from Theorem 4.10. Conversely, consider a vector  $u \in C_f(\bar{x}, \bar{v})$ . Since  $df(\bar{x})$  is the support function of the Fréchet subdifferential  $\hat{\partial}f(\bar{x})$  (see [41, Exercise 8.4]) and by prox-regularity the subdifferentials  $\partial_p f(\bar{x})$  and  $\hat{\partial}f(\bar{x})$  coincide near  $\bar{v}$ , we deduce that  $u$  lies in  $N_{\partial_p f(\bar{x})}(\bar{v})$ . On the other hand by Theorem 4.7, locally near  $\bar{v}$ , we have equality

$$\partial_p f(\bar{x}) = \partial_p(f + \delta_{\mathcal{M}})(\bar{x}) = \nabla g(\bar{x}) + N_{\mathcal{M}}(\bar{x}),$$

where  $g$  is any  $C^2$  smooth function agreeing with  $f$  on  $\mathcal{M}$  near  $\bar{x}$ . Consequently  $u$  lies in  $T_{\mathcal{M}}(\bar{x})$ , as claimed.  $\square$

Next we need set analogues of subderivatives – first-order and second-order tangent sets. These are obtained by applying the subderivative concepts to the indicator function. More concretely we have the following.

**Definition 4.12** (First order and second order tangent sets). Consider a set  $\Omega \subset \mathbf{R}^n$  and a point  $\bar{x} \in \Omega$ . Then the *tangent cone* to  $\Omega$  at  $\bar{x}$  is the set

$$T_{\Omega}(\bar{x}) := \{u : \exists t_i \downarrow 0 \text{ and } u_i \rightarrow u \text{ such that } \bar{x} + t_i u_i \in \Omega\},$$

while the *critical cone of  $\Omega$  at  $\bar{x}$  for  $\bar{v}$*  is defined by

$$C_\Omega(\bar{x}, \bar{v}) := T_\Omega(\bar{x}) \cap \bar{v}^\perp.$$

The *second-order tangent set to  $\Omega$  at  $\bar{x}$  for  $\bar{u} \in T_\Omega(\bar{x})$*  is the set

$$T_\Omega^2(\bar{x}|\bar{u}) := \{w : \exists t_i \downarrow 0 \text{ and } w_i \rightarrow w \quad \text{such that} \quad \bar{x} + t_i \bar{u} + \frac{1}{2} t_i^2 w_i \in \Omega\}.$$

One can now easily verify the relationships:

$$T_\Omega(\bar{x}) = \text{dom } d\delta_Q(\bar{x}), \quad C_\Omega(\bar{x}, \bar{v}) = C_{\delta_Q}(\bar{x}, \bar{v}), \quad T_\Omega^2(\bar{x}|\bar{u}) = \text{dom } d^2\delta_\Omega(\bar{x})(\bar{u}|\cdot).$$

Next we record an important relationship between projections and identifiable manifolds [32, Proposition 4.5]. Naturally, we say that a set  $\mathcal{M}$  is a  *$C^p$  identifiable manifold relative to a set  $Q$  at  $\bar{x}$  for  $\bar{v} \in N_Q(\bar{x})$*  whenever  $\mathcal{M}$  is a  $C^p$  identifiable manifold relative to the indicator function  $\delta_Q$  at  $\bar{x}$  for  $\bar{v} \in \partial\delta_Q(\bar{x})$ .

**Proposition 4.13** (Projections and identifiability). *Consider a closed set  $Q \subset \mathbf{R}^n$  and suppose that  $\mathcal{M}$  is a  $C^p$ -identifiable manifold ( $p \geq 2$ ) at  $\bar{x}$  for  $\bar{v} \in N_Q^p(\bar{x})$ . Then for all sufficiently small  $\lambda > 0$ , the projections  $P_Q$  and  $P_{\mathcal{M}}$  coincide on a neighborhood of  $\bar{x} + \lambda \bar{v}$  and are  $C^{p-1}$ -smooth there.*

**Proposition 4.14** (Second-order tangents to sets with identifiable structure). *Suppose that a closed set  $Q \subset \mathbf{R}^n$  admits an identifiable  $C^3$  manifold at  $\bar{x}$  for  $\bar{v} \in N_Q^p(\bar{x})$ . Consider a nonzero tangent  $\bar{u} \in T_{\mathcal{M}}(\bar{x})$  and a vector  $\bar{w} \in T_Q^2(\bar{x}|\bar{u})$ . Then for any real  $\epsilon > 0$ , there exists  $\hat{u} \in T_{\mathcal{M}}(\bar{x})$  and  $\hat{w} \in T_{\mathcal{M}}^2(\bar{x}|\hat{u})$  satisfying*

$$|\bar{u} - \hat{u}| \leq \epsilon \quad \text{and} \quad \langle \bar{v}, \hat{w} \rangle \geq \langle \bar{v}, \bar{w} \rangle.$$

*Proof.* By definition of  $\bar{w}$ , there exist numbers  $t_i \downarrow 0$  and vectors  $w_i \rightarrow \bar{w}$  so that the points  $x_i := \bar{x} + t_i \bar{u} + \frac{1}{2} t_i^2 w_i$  lie in  $\Omega$  for each  $i$ . By Proposition 4.13, we may choose  $r > 0$  satisfying  $P_Q(\bar{x} + r\bar{v}) = \bar{x}$ , so that  $P_Q$  coincides with  $P_{\mathcal{M}}$  on a neighborhood of  $\bar{x} + r\bar{v}$ , and so that  $P_Q$  is  $C^2$ -smooth on this neighborhood. Define now  $z_i = P_Q(x_i + r\bar{v})$ . Since  $P_Q$  is  $C^2$ -smooth on a neighborhood of  $\bar{x} + r\bar{v}$ , we may write  $z_i = \bar{x} + t_i \hat{u} + \frac{1}{2} t_i^2 \hat{w}_i$  for some  $\hat{u} \in T_{\mathcal{M}}(\bar{x})$  and some  $\hat{w}_i$  converging to a vector  $\hat{w} \in T_{\mathcal{M}}^2(\bar{x}|\hat{u})$ . It is standard that the derivative  $\nabla P_{\mathcal{M}}(\bar{x})$  coincides with the linear projection onto the tangent space  $T_{\mathcal{M}}(\bar{x})$ , and hence decreasing  $r$  we may ensure  $|u - \hat{u}| < \epsilon$ . By definition of  $z_i$  then we have the inequality

$$|x_i - z_i + r\bar{v}| \leq r|\bar{v}|,$$

and hence

$$\langle \bar{v}, z_i - x_i \rangle \geq \frac{1}{2r} |x_i - z_i|^2 \geq 0.$$

We deduce

$$0 \leq \langle \bar{v}, t_i(\hat{u} - \bar{u}) + \frac{1}{2} t_i^2(\hat{w}_i - w_i) \rangle = \frac{1}{2} t_i^2 \langle \bar{v}, \hat{w}_i - \bar{w}_i \rangle.$$

Dividing by  $\frac{1}{2} t_i^2$  and taking the limit the result follows.  $\square$



Finally, we arrive at the key relationship between the parabolic subderivative of a function and that of its restriction to an identifiable manifold.

**Corollary 4.15** (Second-order subderivatives and identifiability). *Suppose that an lsc function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  admits an identifiable  $C^3$  manifold  $\mathcal{M}$  at  $\bar{x}$  for  $0 \in \partial_p f(\bar{x})$ . Consider a nonzero vector  $\bar{u} \in T_{\mathcal{M}}(\bar{x})$  and a vector  $\bar{w}$ . Then for any real  $\epsilon > 0$ , there exists  $\hat{u} \in T_{\mathcal{M}}(\bar{x})$  and  $\hat{w}$  satisfying  $|\bar{u} - \hat{u}| \leq \epsilon$  and*

$$d^2 f(\bar{x})(\bar{u}|\bar{w}) \geq d^2(f + \delta_{\mathcal{M}})(\bar{x})(\hat{u}|\hat{w}).$$

*Proof.* By [16, Proposition 3.14], the set  $\mathcal{K} := \text{gph}(f + \delta_{\mathcal{M}})$  is a  $C^3$  identifiable manifold relative to  $\text{epi } f$  at  $(\bar{x}, f(\bar{x}))$  for  $(\bar{v}, -1)$ . Moreover by Theorem 4.10, we have

$$T_{\mathcal{K}}(\bar{x}) = \{(u, \alpha) : u \in T_{\mathcal{M}}(\bar{x}) \quad \text{and} \quad \alpha = df(\bar{x})(u)\}.$$

Define  $\bar{\beta} := df(\bar{x})(\bar{u})$ . Then by [41, Example 13.62], equality

$$\text{epi } d^2 f(\bar{x})(\bar{u}|\cdot) = T_{\text{epi } f}^2((\bar{x}, f(\bar{x}))|(\bar{u}, \bar{\beta})),$$

holds. Define  $\bar{r} := d^2 f(\bar{x})(\bar{u}|\bar{w})$ . Applying Proposition 4.14, we deduce that there exist  $(\hat{u}, \hat{\beta}) \in T_{\mathcal{K}}(\bar{x}, f(\bar{x}))$  and  $(\hat{w}, \hat{r}) \in T_{\mathcal{K}}^2((\bar{x}, f(\bar{x}))|(\hat{u}, \hat{\beta}))$  satisfying

$$|(\bar{u}, \bar{\beta}) - (\hat{u}, \hat{\beta})| \leq \epsilon \quad \text{and} \quad \langle (0, -1), (\hat{w}, \hat{r}) \rangle \geq \langle (0, -1), (\bar{w}, \bar{r}) \rangle.$$

Clearly  $\hat{\beta} = df(\bar{x})(\hat{u})$  and  $\hat{r} = d^2(f + \delta_{\mathcal{M}})(\bar{x})(\hat{u}|\hat{w})$ . We deduce

$$d^2 f(\bar{x})(\bar{u}|\bar{w}) \geq d^2(f + \delta_{\mathcal{M}})(\bar{x})(\hat{u}|\hat{w}),$$

as claimed.  $\square$

We now arrive at the main result of this section.

**Theorem 4.16** (Generic properties of semi-algebraic problems). *Consider an lsc, semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then there exists an integer  $N > 0$  such that for a generic vector  $v \in \mathbf{R}^n$  the function*

$$f_v(x) := f(x) - \langle v, x \rangle$$

*has no more than  $N$  critical points. Moreover each such critical point  $\bar{x}$  satisfies*

1. **(prox-regularity)**  $f_v$  is prox-regular at  $\bar{x}$  for 0.
2. **(strict complementarity)** The inclusion  $0 \in \text{ri } \partial_p f_v(\bar{x})$  holds.
3. **(identifiable manifold)**  $f_v$  has an identifiable manifold  $\mathcal{M}$  at  $\bar{x}$  for 0.
4. **(smooth dependence of critical points)** The subdifferential  $\partial f$  is strongly regular at  $(\bar{x}, v)$ . More precisely, there exist neighborhoods  $U$  of  $\bar{x}$  and  $V$  of  $v$  so that the critical point mapping

$$w \mapsto U \cap (\partial f)^{-1}(w) = \{x \in U : x \text{ is critical for } f(\cdot) - \langle w, \cdot \rangle\}$$

*is single-valued and analytic on  $V$ , and maps  $V$  onto  $\mathcal{M}$ .*

Moreover the following are all equivalent

- (i)  $\bar{x}$  is a local minimizer of  $f_v$ .
- (ii)  $\bar{x}$  is a stable strong local minimizer of  $f_v$ .
- (iii) The inequality

$$\inf_{w \in \mathbf{R}^n} d^2 f_v(\bar{x})(u|w) > 0 \quad \text{holds for all } 0 \neq u \in C_f(\bar{x}, v).$$

- (iv) The inequality

$$\inf_{w \in \mathbf{R}^n} d^2(f_v + \delta_{\mathcal{M}})(\bar{x})(u|w) > 0 \quad \text{holds for all } 0 \neq u \in T_{\mathcal{M}}(\bar{x}).$$

*Proof.* In light of Corollary 4.8, we must only argue the claimed equivalence of the four properties. To this end, observe that for generic  $v$ , the equivalence (i)  $\Leftrightarrow$  (ii) was established in Corollary 4.8. On the other hand, Theorem 4.7 shows that (ii) is equivalent to  $\bar{x}$  being a strong local minimizer of  $f_v$  on  $\mathcal{M}$ , which in turn for classical reasons is equivalent to (iv). Note also that the implication (iii)  $\Rightarrow$  (iv) is obvious from Theorem 4.11. Thus we must only show the implication (iv)  $\Rightarrow$  (iii), but this follows immediately from Corollary 4.15.  $\square$

Note that property (iv) in the theorem above involves only classical analysis.

## 5 Composite semi-algebraic optimization

In this section, we consider composite optimization problems of the form

$$\min f(x) + h(G(x)),$$

where  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and  $h: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$  are lsc functions and  $G: \mathbf{R}^n \rightarrow \mathbf{R}$  is  $C^2$ -smooth. A prime example is the case of smoothly constrained optimization; this is the case where  $h$  is the indicator function of a closed set. We call a point  $x \in \mathbf{R}^n$  *composite critical* for the problem if there exists a vector

$$\lambda \in \partial h(G(x)) \quad \text{satisfying} \quad -\nabla G(x)^* \lambda \in \partial f(x).$$

Whenever the optimality condition above holds, we call  $\lambda$  a *Lagrange multiplier vector* and the tuple  $(x, \lambda)$  a *composite critical pair*. The multiplier  $\lambda$  is sure to be unique under the condition:

$$(1) \quad \text{par } \partial h(G(x)) \bigcap [\nabla G(x)^*]^{-1} \text{par } \partial f(x) = \{0\}.$$

Indeed, this is a direct analogue of the linear independence constraint qualification in nonlinear programming.

In general, the notion of composite criticality is different from criticality (as defined in the previous sections) for the function  $f + h \circ G$ . If  $x$  is a critical point of  $f + h \circ G$ , then  $x$  is composite critical only under some additional condition, such as the *basic constraint qualification*

$$(2) \quad \partial^\infty h(G(x)) \bigcap [\nabla G(x)^*]^{-1} \partial^\infty f(x) = \{0\}.$$

This qualification condition is a generalization of the Mangasarian-Fromovitz constraint qualification in nonlinear programming and is in particular implied by (1); see the discussion in [39] for more details. Conversely, if  $x$  is a composite critical point and both  $f$  and  $h$  are subdifferentially regular [41, Definition 7.25] (as is the case when  $f$  and  $h$  are convex), then  $x$  is also a critical point of the function  $f + h \circ G$ .

In this section, we consider properties of composite critical points for generic composite semi-algebraic problems. To this end, we will assume that  $f$ ,  $G$ , and  $h$  are all semi-algebraic and we will consider the canonically perturbed problems:

$$\min f(x) + h(G(x) + y) - \langle v, x \rangle.$$

Then composite criticality is succinctly captured by the generalized equation

$$(3) \quad \begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left( \partial f \times (\partial h)^{-1} \right)(x, \lambda).$$

The path to generic properties is now clear since the perturbation parameters  $(v, y)$  appear in the range space of a semi-algebraic set-valued mapping having a small graph.

Before we proceed, we briefly recall that subderivatives admit a convenient calculus [41, Exercise 13.63] for the composite problem. In what follows, for any  $C^2$ -smooth mapping  $G(x) = (g_1(x), \dots, g_m(x))$  we use the notation

$$\nabla^2 G(x)[u, u] = (\langle \nabla^2 g_1(x)u, u \rangle, \dots, \langle \nabla^2 g_m(x)u, u \rangle).$$

**Theorem 5.1** (Calculus of subderivatives).

*Consider a  $C^2$ -smooth mapping  $G: \mathbf{R}^n \rightarrow \mathbf{R}^m$  and lsc functions  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and  $h: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ . Suppose that a point  $x$  satisfies the constraint qualification*

$$\partial^\infty h(G(x)) \cap [\nabla G(x)^*]^{-1} \partial^\infty f(x) = \{0\}.$$

*Then the equality*

$$d(f + h \circ G)(x)(u) = df(x)(u) + dh(G(x))(\nabla G(x)u)$$

*holds. Moreover for any  $u$  with  $d(f + h \circ G)(x)(u)$  finite, we have*

$$d^2(f + h \circ G)(x)(u|w) = d^2 f(x)(u|w) + d^2 h(G(x)) \left( \nabla G(x)u \mid \nabla^2 G[u, u] + \nabla G(x)w \right).$$

We are now ready to prove the main result of this section. Note that if for almost every  $v$ , a property is valid for almost every  $y$  (with the  $v$  fixed), then by Fubini's theorem the said property holds for almost every pair  $(v, y)$ . The same holds with  $v$  and  $y$  reversed. We will use this observation implicitly throughout.

**Theorem 5.2** (Generic properties of composite optimization problems). *Consider a  $C^2$ -smooth semi-algebraic mapping  $G: \mathbf{R}^n \rightarrow \mathbf{R}^m$  and lsc semi-algebraic functions  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and  $h: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ . Define now the family of composite optimization problems  $P(v, y)$  by*

$$\min f_v(x) + h(G_y(x)),$$

under the perturbations  $f_v(x) := f(x) - \langle v, x \rangle$  and  $G_y(x) = G(x) + y$ . Then for almost every  $y \in \mathbf{R}^m$ , the qualification conditions

$$(4) \quad \text{span } \partial^\infty h(G_y(x)) \bigcap [\nabla G(x)^*]^{-1} \text{span } \partial^\infty f_v(x) = \{0\},$$

$$(5) \quad \text{par } \partial h(G_y(x)) \bigcap [\nabla G(x)^*]^{-1} \text{par } \partial f_v(x) \subseteq \{0\},$$

hold for any  $x$ , for which  $f_v(x)$  and  $h(G_y(x))$  are finite. Moreover there exists an integer  $N > 0$  such that for a generic collection of parameters  $(v, y) \in \mathbf{R}^n \times \mathbf{R}^m$ , the problem  $P(v, y)$  has at most  $N$  composite critical points, and for any such composite critical point  $\bar{x}$  of  $P(v, y)$ , there exist a unique Lagrange multiplier vector

$$\bar{\lambda} \in \partial h(G_y(\bar{x})) \quad \text{satisfying} \quad -\nabla G(\bar{x})^* \bar{\lambda} \in \partial f_v(\bar{x}).$$

Moreover, defining  $\bar{w} := -\nabla G(\bar{x})^* \bar{\lambda}$ , the following are true.

1. **(prox-regularity)**  $f_v$  is prox-regular at  $\bar{x}$  for  $\bar{w}$  and  $h$  is prox-regular at  $G_y(\bar{x})$  for  $\bar{\lambda}$ .
2. **(strict-complementarity)** The inclusions

$$\bar{\lambda} \in \text{ri } \partial_p h(G_y(\bar{x})) \quad \text{and} \quad \bar{w} \in \text{ri } \partial_p f_v(\bar{x}) \quad \text{hold.}$$

3. **(identifiable manifold)**  $f_v$  admits a  $C^\omega$  identifiable manifold  $\mathcal{M}$  at  $\bar{x}$  for  $\bar{w}$  and  $h$  admits a  $C^\omega$  identifiable manifold  $\mathcal{K}$  at  $G_y(\bar{x})$  for  $\bar{\lambda}$ .
4. **(nondegeneracy)** The constraint qualification (nondegeneracy condition)

$$N_{\mathcal{K}}(G_y(\bar{x})) \cap [\nabla G(\bar{x})^*]^{-1} N_{\mathcal{M}}(\bar{x}) = \{0\} \quad \text{holds.}$$

5. **(smooth dependence of critical triples)** The mapping

$$(\hat{v}, \hat{y}) \mapsto \left\{ (x, \lambda) : \text{the pair } (x, \lambda) \text{ is composite critical for } P(\hat{v}, \hat{y}) \right\}$$

admits a single-valued analytic localization around  $(v, y, \bar{x}, \bar{\lambda})$ .

Moreover the following are equivalent.

- (i)  $\bar{x}$  is a local minimizer of  $P(v, y)$ .
- (ii)  $\bar{x}$  is a strong local minimizer of  $P(v, y)$ .
- (iii) The inequality

$$d^2 f(\bar{x})(u|z) + d^2 h(G_y(\bar{x})) \left( \nabla G(\bar{x}) u \mid \nabla^2 G(\bar{x})[u, u] + \nabla G(\bar{x}) z \right) > 0$$

holds for all nonzero  $u \in C_f(\bar{x}, \bar{w}) \cap [\nabla G(\bar{x})]^{-1} C_h(G_y(\bar{x}), \bar{\lambda})$  and all  $z \in \mathbf{R}^n$ .

- (iv) The inequality

$$d^2 (f + \delta_{\mathcal{M}})(\bar{x})(u|z) + d^2 (h + \delta_{\mathcal{K}})(G_y(\bar{x})) \left( \nabla G(\bar{x}) u \mid \nabla^2 G(\bar{x})[u, u] + \nabla G(\bar{x}) z \right) > 0$$

holds for all nonzero  $u \in T_{\mathcal{M}}(\bar{x}) \cap [\nabla G(\bar{x})]^{-1} T_{\mathcal{K}}(G_y(\bar{x}))$  and all  $z \in \mathbf{R}^n$ .

*Proof.* First applying [4, Lemma 8] and Theorem 3.4, we obtain a  $C^\omega$  stratification  $\{A_i\}$  of  $\text{dom } f$  and a  $C^\omega$  stratification  $\{B_j\}$  of  $\text{dom } h$  having the property that  $f$  is  $C^\omega$ -smooth on each  $A_i$  and  $h$  is  $C^\omega$ -smooth on each  $B_j$ , and so that

$$\partial^\infty f(x) \cup \text{par } \partial f(x) \subset N_{A_i}(x) \quad \text{and} \quad \partial^\infty h(z) \cup \text{par } \partial h(z) \subset N_{B_j}(z)$$

for any  $x \in A_i$  and  $z \in B_j$ . For fixed indices  $i$  and  $j$ , the standard Sard's theorem implies that for almost every  $y \in \mathbf{R}^m$ , the restriction of  $G_y$  to  $A_i$  is transverse to  $B_j$ , that is for any  $x \in A_i$  with  $G_y(x) \in B_j$  we have

$$N_{B_j}(G_y(x)) \cap [\nabla G(x)^*]^{-1} N_{A_i}(x) = \{0\}.$$

Since there are finitely many indices  $i$  and  $j$ , the claimed qualification conditions (4) and (5) follow.

Define now the set-valued mapping  $\mathcal{I}: \mathbf{R}^n \times \mathbf{R}^m \rightrightarrows \mathbf{R}^n \times \mathbf{R}^m$  by

$$\mathcal{I}(x, \lambda) = \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left( \partial f \times (\partial h)^{-1} \right)(x, \lambda).$$

Observe  $(v, y) \in \mathcal{I}(x, \lambda)$  if and only if  $(x, \lambda)$  is a composite critical pair for  $P(v, y)$ . It is easy to see, in turn, that  $\text{gph } \mathcal{I}$  is  $C^1$  diffeomorphic to  $\text{gph } \partial f \times \text{gph } (\partial h)^{-1}$ , and hence by [14, Theorem 3.7] has dimension  $n + m$ . Applying the semi-algebraic Sard's theorem for weakly critical values (Theorem 3.7), we deduce that there exists an integer  $N > 0$  such that for generic parameters  $(v, y)$ , the problem  $P(v, y)$  has at most  $N$  composite critical points  $x$ . Moreover for any composite critical point  $\bar{x}$  of  $P(v, y)$ , the Lagrange multiplier vector  $\bar{\lambda}$  is unique for almost every  $(v, y)$  by inclusion (5).

We now prove the strict complementarity claim. To this end, define the mapping

$$\mathcal{I}_p(x, \lambda) = \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left( \text{ri } \partial_p f \times (\text{ri } \partial_p h)^{-1} \right)(x, \lambda).$$

Clearly the inclusion  $\text{gph } \mathcal{I}_p \subset \text{gph } \mathcal{I}$  holds, and by what we have already proved both mappings  $\mathcal{I}_p$  and  $\mathcal{I}$  are finite valued almost everywhere. We now claim that  $\text{gph } \mathcal{I}_p$  is dense in  $\text{gph } \mathcal{I}$ . To see this, fix a pair  $(v, y) \in \mathcal{I}(x, \lambda)$ . Equivalently we may write

$$0 = w + \nabla G(x)^* \lambda, \quad \text{for some } w \in \partial f_v(x) \text{ and } \lambda \in \partial h(G_y(x)).$$

By definition of the limiting subdifferential, there are sequences  $(x_k, u_k) \rightarrow (x, w + v)$  in  $\text{gph } (\text{ri } \partial_p f)$  and  $(z_k, \lambda_k) \rightarrow (G_y(x), \lambda)$  in  $\text{gph } (\text{ri } \partial_p h)$ . Defining  $\gamma_k := (u_k - (w + v)) + (\nabla G(x_k)^* \lambda_k - \nabla G(x)^* \lambda)$  and  $\alpha_k := z_k - G_y(x_k)$  it is easy to verify the inclusion

$$(v + \gamma_k, y + \alpha_k) \in \mathcal{I}_p(x_k, \lambda_k).$$

Hence  $\text{gph } \mathcal{I}_p$  is dense in  $\text{gph } \mathcal{I}$ . Since both  $\mathcal{I}^{-1}$  and  $\mathcal{I}_p^{-1}$  are semi-algebraic and finite almost everywhere, it follows immediately that  $\mathcal{I}^{-1}$  and  $\mathcal{I}_p^{-1}$  agree almost everywhere on  $\mathbf{R}^n \times \mathbf{R}^m$ . This establishes the strict complementarity claim 2.

Moving on to existence of identifiable manifolds, applying Theorem 3.7 to the mapping  $\mathcal{I}_p$ , we deduce that there exists an integer  $N$ , a finite collection of open

semi-algebraic sets  $\{U_i\}_{i=0}^N$  in  $\mathbf{R}^n \times \mathbf{R}^m$ , and analytic semi-algebraic single-valued mappings

$$E_i^j : U_i \rightarrow \mathbf{R}^n \times \mathbf{R}^m \quad \text{for } i = 0, \dots, N \text{ and } j = 1, \dots, i$$

satisfying:

1.  $\bigcup_i U_i$  is dense in  $\mathbf{R}^n \times \mathbf{R}^m$ .
2. For any  $(v, y) \in U_i$ , the image  $\mathcal{I}_p^{-1}(v, y)$  has cardinality  $i$ .
3. We have the representation

$$\mathcal{I}_p^{-1}(v, y) = \{E_i^j(v, y) : j = 1, 2, \dots, i\} \quad \text{whenever } (v, y) \in U_i.$$

Let  $X_i^j(v, y)$  denote the composition of  $E_i^j$  with the projection  $(x, \lambda) \mapsto x$  and let  $F_i^j(v, y) := G(X_i^j(v, y)) + y$ . Applying Theorem 3.4 to each  $X_i^j$  and  $F_i^j$ , we may find a dense open subset  $\widehat{U}_i$  of  $U_i$  so that

- $f$  is analytic on  $X_i^j(\widehat{U}_i)$  and  $h$  is analytic on  $F_i^j(\widehat{U}_i)$ .
- $X_i^j$  and  $F_i^j$  are analytic and have constant rank on  $\widehat{U}_i$

Let  $(x, \lambda)$  be such that  $X_i^j(v, y) = x$  and so that  $(x, \lambda)$  is a composite critical pair for  $P(v, y)$ . Define also  $w := -\nabla G(x)^* \lambda$ . Then due to the constant rank, there exists a neighborhood  $W$  of  $(v, y)$  so that  $X_i^j(W)$  and  $F_i^j(W)$  are analytic manifolds. We claim that  $X_i^j(W)$  is an identifiable manifold relative to  $f_v$  at  $x$  for  $w$  and that  $F_i^j(W)$  is an identifiable manifold relative to  $h$  at  $G_y(x)$  for  $\lambda$ .

To see this, consider sequences  $(x_k, w_k) \rightarrow (x, w)$  in  $\text{gph } \partial f_v$  and  $(z_k, \lambda_k) \rightarrow (G_y(x), \lambda)$  in  $\text{gph } \partial h$ . Defining  $\gamma_k := (w_k - w) + (\nabla G(x_k)^* \lambda_k - \nabla G(x)^* \lambda)$  and  $\alpha_k := z_k - G_y(x_k)$  we have the inclusion

$$(v + \gamma_k, y + \alpha_k) \in \mathcal{I}_p(x_k, \lambda_k).$$

Hence for all large indices  $k$  equality

$$E_j^i(v + \gamma_k, y + \alpha_k) = (x_k, \lambda_k)$$

holds. We deduce for sufficiently large  $k$  the inclusion  $x_k \in X_i^j(W)$ . Hence  $X_i^j(W)$  is indeed identifiable relative to  $f_v$  at  $x$  for  $w$ . Moreover, we have  $z_k = F_i^j(v + \gamma_k, y + \alpha_k) \in F_i^j(W)$  for all large  $k$ . We conclude that  $F_i^j(W)$  is identifiable relative to  $h$  at  $G_y(x)$  for  $\lambda$ , as claimed. The nondegeneracy claim is a simple consequence of the construction and the classical Sard's theorem. Finally the four equivalent properties are immediate from Theorems 4.16 and 5.1.  $\square$

Note that Theorem 5.2 with  $h = 0$  and  $G = I$  reduces to Theorem 4.16. It is interesting to reinterpret Theorem 5.2 in the convex setting. To this end, recall that for any convex function  $f : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , the *Fenchel conjugate*  $f^* : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is defined by

$$f^*(u) := \sup_x \{\langle u, x \rangle - f(x)\},$$

and the relationship  $\partial f^* = (\partial f)^{-1}$  holds.

Fix now a linear mapping  $A: \mathbf{R}^n \rightarrow \mathbf{R}^m$  and lsc convex functions  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and  $h: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ . Within the Fenchel framework, we consider the family of primal optimization problems given by

$$\inf_x f(x) + h(Ax + y) - \langle v, x \rangle,$$

and associate with them the dual problems

$$\sup_u -h^*(u) - f^*(v - A^*u) + \langle y, u \rangle.$$

Then the primal problem is feasible whenever  $y$  lies in the set

$$Y := \text{dom } h - A(\text{dom } f),$$

and the dual is feasible whenever  $v$  lies in

$$V := \text{dom } f^* + A^*(\text{dom } h^*).$$

Standard Fenchel duality then asserts that for  $y$  in the interior of  $Y$ , the primal and dual optimal values are equal and the dual is attained when finite. Assuming in addition that  $v$  lies in the interior of  $V$ , optimality is characterized by the generalized equation

$$\begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} A^*u \\ -Ax \end{bmatrix} + (\partial f \times \partial h^*)(x, u).$$

This is precisely an instance of the variational inequality (3) in a convex setting. Assuming now that  $f$  and  $h$  are semi-algebraic, and applying Theorem 5.2, we deduce that for generic parameters  $(v, y)$ , if the primal and dual problems are feasible then the interiority conditions hold, and both the primal and the dual admit at most one minimizer. Moreover for any such minimizers  $x$  and  $u$ , strict complementarity holds for the primal and the dual, identifiable manifolds exist for both problems, both objectives grow quadratically around  $x$  and  $u$ , respectively, and the minimizers  $x$  and  $u$  jointly vary analytically with the parameters  $(v, y)$ .

## References

- [1] F. Al-Khayyal and J. Kyparisis. Finite convergence of algorithms for nonlinear programs and variational inequalities. *J. Optim. Theory Appl.*, 70(2):319–332, 1991.
- [2] F. Alizadeh, J.-P.A. Haeberly, and M.L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, 77(2, Ser. B):111–128, 1997. Semidefinite programming.
- [3] J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semi-algebraic convex programs. *Math. Oper. Res.*, 36:55–70, 2011.
- [4] J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optimization*, 18(2):556–572, 2007.



- [5] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000.
- [6] J.V. Burke. On the identification of active constraints. II. The nonconvex case. *SIAM J. Numer. Anal.*, 27(4):1081–1103, 1990.
- [7] J.V. Burke and J.J. Moré. On the identification of active constraints. *SIAM J. Numer. Anal.*, 25(5):1197–1211, 1988.
- [8] P.H. Calamai and J.J. Moré. Projected gradient methods for linearly constrained problems. *Math. Program.*, 39(1):93–116, 1987.
- [9] M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.
- [10] D. Drusvyatskiy and A.D. Ioffe. Quadratic growth and critical point stability of semi-algebraic functions. *To appear in Math. Program.*, *arXiv:1309.1446*, 2014.
- [11] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. The dimension of semialgebraic subdifferential graphs. *Nonlinear Anal.*, 75(3):1231–1245, 2012.
- [12] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Clarke subgradients for directionally lipschitzian stratifiable functions. *Math. Oper. Res.*, 40(2):328–349, 2015.
- [13] D. Drusvyatskiy and A.S. Lewis. Generic nondegeneracy in convex optimization. *Proc. Amer. Math. Soc.*, 139(7):2519–2527, 2011.
- [14] D. Drusvyatskiy and A.S. Lewis. Semi-algebraic functions have small subdifferentials. *Math. Program.*, 140(1, Ser. B):5–29, 2013.
- [15] D. Drusvyatskiy and A.S. Lewis. Tilt stability, uniform quadratic growth, and strong metric regularity of the subdifferential. *SIAM J. Optim.*, 23(1):256–267, 2013.
- [16] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. *Math. Program.*, 147(1-2, Ser. A):467–498, 2014.
- [17] D. Drusvyatskiy, B.S. Mordukhovich, and T.T.A. Nghia. Second-order growth, tilt stability, and metric regularity of the subdifferential. *J. Convex Anal.*, 21(4):1165–1192, 2014.
- [18] J.C. Dunn. On the convergence of projected gradient processes to singular critical points. *J. Optim. Theory Appl.*, 55(2):203–216, 1987.
- [19] F. Facchinei and J. Pang. *Finite dimensional variational inequalities and complementarity problems*. Springer Series in Operations Research, Springer-Verlag, New York, 2003.
- [20] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems. Vol. II*. Springer Series in Operations Research. Springer-Verlag, New York, 2003.
- [21] M.C. Ferris. Finite termination of the proximal point algorithm. *Math. Program. Ser. A*, 50(3):359–366, 1991.

- [22] S.D. Flåm. On finite convergence and constraint identification of subgradient projection methods. *Math. Program.*, 57:427–437, 1992.
- [23] W.L. Hare and A.S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *J. Convex Anal.*, 11(2):251–266, 2004.
- [24] W.L. Hare and A.S. Lewis. Identifying active manifolds. *Algorithmic Oper. Res.*, 2(2):75–82, 2007.
- [25] A.D. Ioffe. Metric regularity and subdifferential calculus. *Uspekhi Mat. Nauk*, 55(3(333)):103–162, 2000.
- [26] A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. extensions of Sard and Smale-Sard theorems. *Proc. Amer. Math. Soc.*, 136:3111 – 3119, 2008.
- [27] A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19(4):1894–1917, 2009.
- [28] J.M. Lee. *Introduction to Smooth Manifolds*. Springer, New York, 2003.
- [29] S. Lee and S.J. Wright. Manifold identification in dual averaging for regularized stochastic online learning. *J. Mach. Learn. Res.*, 13:1705–1744, 2012.
- [30] A.S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13:702–725, 2002.
- [31] A.S. Lewis. Nonsmooth optimization: conditioning, convergence and semi-algebraic models. In *Proceedings of the International Congress of Mathematicians, Seoul*, pages 871–895, 2014.
- [32] A.S. Lewis and S. Zhang. Partial smoothness, tilt stability, and generalized Hessians. *SIAM J. Optim.*, 23(1):74–94, 2013.
- [33] R. Mifflin and C. Sagastizábal. A  $\mathcal{VU}$ -algorithm for convex minimization. *Math. Program.*, 104(2-3, Ser. B):583–608, 2005.
- [34] G.J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29:341–346, 1962.
- [35] G. Pataki and L. Tunçel. On the generic properties of convex optimization problems in conic form. *Math. Program.*, 89(3, Ser. A):449–457, 2001.
- [36] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.
- [37] S.M. Robinson. A point-of-attraction result for Newton’s method with point-based approximations. *Optimization*, 60(1-2):89–99, 2011.
- [38] S.M. Robinson. Equations on monotone graphs. *Math. Program.*, 141(1-2, Ser. A):49–101, 2013.
- [39] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Rev.*, 35(2):183–238, 1993.
- [40] R.T. Rockafellar and A.L. Dontchev. *Implicit functions and solution mappings*. Monographs in Mathematics, Springer-Verlag, 2009.
- [41] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.

- [42] R. Saigal and C. Simon. Generic properties of the complementarity problem. *Math. Programming*, 4:324–335, 1973.
- [43] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Math. Programming*, 77(2, Ser. B):301–320, 1997. Semidefinite programming.
- [44] J.E. Spingarn. On optimality conditions for structured families of nonlinear programming problems. *Math. Programming*, 22(1):82–92, 1982.
- [45] J.E. Spingarn and R.T. Rockafellar. The generic nature of optimality conditions in nonlinear programming. *Mathematics of Operations Research*, 4(4):pp. 425–430, 1979.
- [46] L. Takács. An increasing continuous singular function. *Amer. Math. Monthly*, 85(1):35–37, 1978.
- [47] L. van den Dries and C. Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84:497–540, 1996.
- [48] S.J. Wright. Identifiable surfaces in constrained optimization. *SIAM J. Control Optim.*, 31:1063–1079, July 1993.